

# Response-adaptive dose-finding under model uncertainty

Björn Bornkamp

Fakultät Statistik

Technische Universität Dortmund

44221 Dortmund, Germany

e-mail: bornkamp@statistik.tu-dortmund.de

Frank Bretz

Statistical Methodology

Novartis Pharma AG

4002 Basel, Switzerland

e-mail: frank.bretz@novartis.com

Holger Dette

Fakultät für Mathematik

Ruhr-Universität Bochum

44780 Bochum, Germany

e-mail: holger.dette@ruhr-uni-bochum.de

José Pinheiro

Johnson & Johnson

Raritan, NJ 08869, USA

e-mail: jpinheil@its.jnj.com

April 22, 2010

## Abstract

Dose-finding studies are frequently conducted to evaluate the effect of different doses or concentration levels of a compound on a response of interest. Applications include the investigation of a new medicinal drug, a herbicide or fertilizer, a molecular entity, an environmental toxin, or an industrial chemical. In pharmaceutical drug development, dose-finding studies are of critical importance because of regulatory requirements that marketed doses are safe and provide clinically relevant efficacy. Motivated by a dose-finding study in moderate persistent asthma, we propose response-adaptive designs addressing two major

challenges in dose-finding studies: uncertainty about the dose-response models and large variability in parameter estimates. To allocate new cohorts of patients in an ongoing study, we use optimal designs that are robust under model uncertainty. In addition, we use a Bayesian shrinkage approach to stabilize the parameter estimates over the successive interim analyses used in the adaptations. This allows us to calculate updated parameter estimates and model probabilities that can then be used to calculate the optimal design for subsequent cohorts. The resulting designs are hence robust with respect to model misspecification and additionally can efficiently adapt to the information accrued in an ongoing study. We focus on adaptive designs for estimating the minimum effective dose, although alternative optimality criteria or mixtures thereof could be used, enabling the design to address multiple objectives. In an extensive simulation study, we investigate the operating characteristics of the proposed methods under a variety of scenarios discussed by the clinical team to design the aforementioned clinical study.

*Keywords:* Dose-response; Drug development; Minimum effective dose; Optimal design; Shrinkage approach.

## 1 Introduction

Dose-finding studies have several challenges in common. First, they usually address two distinct objectives, which lead to different requirements on the study design (Ruberg (1995); Bretz, Hsu, Pinheiro & Liu (2008)): (i) assessing whether there is any evidence of a drug effect, and (ii) estimating relevant target doses. Second, the dose-response relationship is unknown prior to the study, leading to model uncertainty. This problem is often underestimated, although ignoring model uncertainty can lead to highly undesirable effects (Chatfield (1995), Draper (1995), Hjorth (1994)). Third, data from dose-finding studies are usually highly variable. This of particular importance in pharmaceutical drug development, because sample sizes are kept to a minimum for ethical and financial reasons. It is therefore critical to develop efficient dose-finding study designs that use the limited information as efficiently as possible, while addressing the above challenges.

Many approaches have been proposed in the optimal design literature to distribute patients efficiently with regard to given study objectives, see Wu (1988), Fedorov & Leonov (2001) and King & Wong (2004) among many others. However, most of this work has concentrated on an assumed fixed dose-response model. As there is typically considerable model uncertainty at the planning stage of a dose-response study, these methods have limited practical use. Based on concepts introduced by Läuter (1974) (see also Cook & Wong (1994), Zhu & Wong (2000), Zhu & Wong (2001), Biedermann, Dette & Pepelyshev (2006)), Dette, Bretz, Pepelyshev & Pinheiro (2008) investigated model-robust designs that provide efficient target dose estimates for a set of candidate dose-response models, rather than for a single dose-response model. However, their designs require knowledge about the unknown parameters associated with the anticipated dose-response models as well as the pre-specification of model probabilities.

A natural remedy is to investigate response-adaptive designs (adaptive designs, in short) with several cohorts of subjects. After each stage the accumulated information of the ongoing study is used to update the initial information of the underlying model parameters and model probabilities, which in turn is used to calculate the design for the subsequent stage(s). Several adaptive designs have been developed for this problem; see, for example, Miller, Guilbaud & Dette (2007) and Dragalin, Hsuan & Padmanabhan (2007) for recent approaches using optimal design theory, or Zhou, Joseph, Wolfson & Bélisle (2003), Müller, Berry, Grieve & Krams (2006), and Wathen & Thall (2008) for recent Bayesian adaptive designs. Dragalin, Bornkamp, Bretz, Miller, Padmanabhan, Patel, Perevozskaya, Pinheiro & Smith (2010) performed an extensive simulation study that compared five different adaptive dose-finding methods.

In this paper we propose adaptive designs addressing the three major challenges described above: multiple study objectives, model uncertainty and large variability in the data. For this purpose we use the model-robust designs proposed by Dette et al. (2008) together with a Bayesian shrinkage approach to stabilize the parameter estimates, especially in the early part of a study. This allows one to calculate parameter estimates as well as model probabilities that can then be used to calculate model-robust designs for the subsequent stage(s) of the study. The resulting designs are robust with respect to model misspecification and additionally adapt to the continuously

accrued information in an ongoing study. We focus on adaptive designs for estimating the minimum effective dose (MED), i.e. the smallest dose achieving a clinically relevant benefit over the placebo response. However, alternative optimality criteria or mixtures of optimality criteria could be used, enabling the design to address multiple objectives.

## 2 Asthma dose-finding study

The research for this article was motivated by a Phase II dose-finding study for the development of a new pharmaceutical compound in asthma. This was a multi-center, randomized, double-blind, placebo controlled, parallel group study in patients with moderate persistent asthma, who were randomized to one of seven active dose levels or placebo. The primary endpoint was a lung function parameter (forced expiratory volume in 1 second,  $FEV_1$ ) measured after 28 days of administration, scaled such that larger values indicated a better outcome. The objective of the trial was to evaluate the dose effects over placebo for the primary endpoint and to assess whether there was any evidence of a drug effect. Once such a dose-response signal had been detected, one would subsequently estimate relevant target doses, where the primary focus was on estimating the MED.

Based on discussions with the clinical team, a homoscedastic normal model was assumed for the primary endpoint with a standard deviation of 350 mL, a placebo effect of 100 mL and a maximum treatment effect of 300 mL within the dose range  $[0, 50]$  under investigation. The available doses were 0 (= placebo), 0.5, 1.0, 2.5, 5, 10, 20, and 50. The clinically relevant benefit over the placebo effect was set to 200 mL. That is, an increase in treatment effect of less than 200 mL over the observed placebo response was considered to be clinically irrelevant. Furthermore, all dose levels within the investigated dose range were considered safe based on previous studies, so that efficacy was of primary interest.

Because this study was conducted early in the drug development program, limited information about the dose-response shape was available at the planning stage. A set of candidate dose-response models was derived before starting the study, see Table 1 and Figure 1 for the full model

specifications (including a preliminary specification of the model parameters). Two concave increasing models ( $\text{Emax}_1$ ,  $\text{Emax}_2$ ) were included to describe an increase of the dose-response curve in the lower part of the investigated dose range. In addition, S-shaped ( $\text{Logistic}_1$ ), unimodal (Beta) and convex ( $\text{Logistic}_2$ ) models were included in the candidate model set to robustify the statistical analysis with respect to model uncertainty. We refer to Pinheiro, Bornkamp & Bretz (2006) for details on the use of candidate models in dose-response studies and the elicitation of best guesses for the model parameters.

| Model               | Full model specification  | Model parameters      | True MED |
|---------------------|---|-----------------------|----------|
| Beta                | $\theta_0 + \theta_1 B(\theta_2, \theta_3)(d/60)^{\theta_2}(1 - d/60)^{\theta_3}$ | (100, 300, 0.43, 0.6) | 5.21     |
| $\text{Emax}_1$     | $\theta_0 + \theta_1 d / (\theta_2 + d)$  | (100, 420, 20)        | 18.18    |
| $\text{Emax}_2$     | $\theta_0 + \theta_1 d / (\theta_2 + d)$  | (100, 330, 5)         | 7.69     |
| $\text{Logistic}_1$ | $\theta_0 + \theta_1 / \{1 + \exp[(\theta_2 - d)/\theta_3]\}$                     | (98, 302, 17.5, 3.3)  | 19.82    |
| $\text{Logistic}_2$ | $\theta_0 + \theta_1 / \{1 + \exp[(\theta_2 - d)/\theta_3]\}$                     | (92, 615, 50, 11.5)   | 42.28    |

Table 1: *Candidate dose response models as a function of dose  $d$ , where  $B(a, b) = (a + b)^{a+b} / (a^a b^b)$ .*

Given the information and constraints above, the clinical team was faced at the planning stage with several remaining key questions on the study design:

- (A) Should an adaptive design be employed at all or would a non-adaptive design be sufficient?
- (B) If the decision was to employ an adaptive design, how many interim analyses should be conducted?
- (C) How many dose levels should be included in the study, i.e. are all seven active dose levels from above needed?
- (D) If not all active dose levels were needed, which of them should then be investigated?

In addition to these statistical questions, many further considerations were discussed by the clinical team: adaptive designs require more logistical effort to set-up the repeated data collection

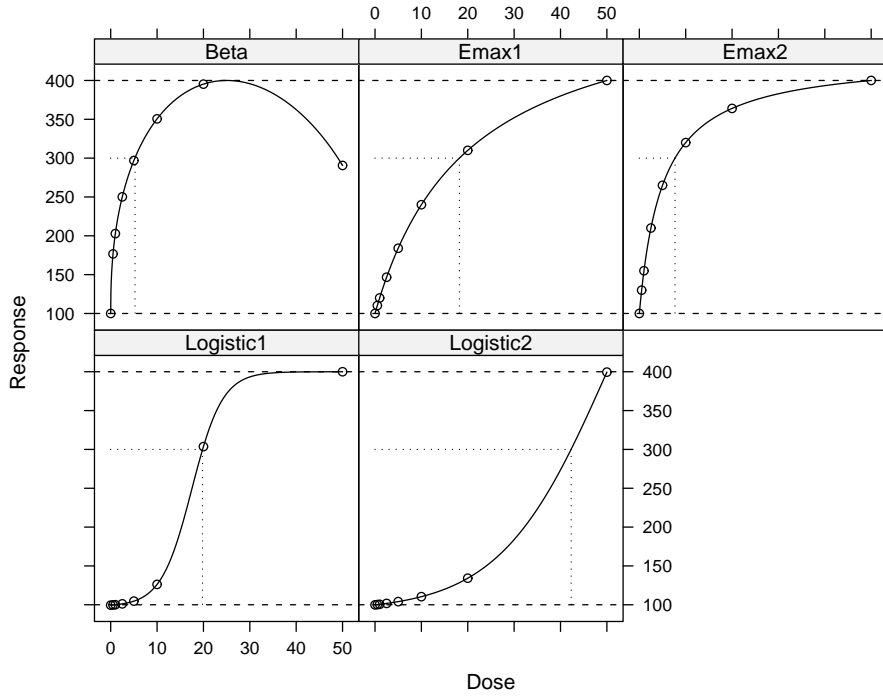


Figure 1: *Graphical display of the dose-response models in Table 1. Open dots denote the potential responses at the seven active dose levels and placebo available in the study. Dotted horizontal lines indicate the clinical relevance threshold on top of placebo response and dotted vertical lines the resulting MED.*

and cleaning/analysis processes than non-adaptive designs; including all seven active doses in the study would pose serious challenges to the drug manufacturing and supply departments, especially if the allocation changed during the study course; and how to ensure trial integrity and validity. In the following we focus on the statistical questions and describe the proposed methodology that was finally implemented in the study.

### 3 Methodology

Assume  $k$  distinct dose levels  $d_1, \dots, d_k$ , where  $d_1 = 0$  denotes placebo. Let  $n_i$  patients be allocated to dose  $d_i$  and  $N = \sum_{i=1}^k n_i$ . The vector of allocation weights is denoted by  $\mathbf{w} = (w_1, \dots, w_k)'$ , where  $w_i = n_i/N$ . Let further  $Y_{ij} \sim N(f(d_i, \boldsymbol{\theta}), \sigma^2)$  denote the observation of patient  $j = 1, \dots, n_i$  at dose  $d_i, i = 1, \dots, k$ , where the dose-response model  $f(\cdot)$  is parameterized through the parameter vector  $\boldsymbol{\theta}$  and  $N(\mu, \sigma^2)$  denotes the normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

Most dose-response models used in practice, including those in Table 1, can be decomposed as

$$f(d, \boldsymbol{\theta}) = \theta_0 + \theta_1 f^0(d, \boldsymbol{\theta}^0), \quad (1)$$

where  $\boldsymbol{\theta} = (\theta_0, \theta_1, \boldsymbol{\theta}^0)' = (\theta_0, \dots, \theta_p)'$ . The parameters  $\boldsymbol{\theta}^* = (\theta_0, \theta_1)'$  enter the model function  $f$  linearly and determine its location and scale, while  $f^0$  is typically a non-linear function that determines the shape of the model function  $f$  through the parameters  $\boldsymbol{\theta}^0$ .

The minimum effective dose producing a clinically relevant effect  $\Delta$  over the placebo response is defined as

$$\text{MED} = \min_{d \in (d_1, d_k]} \{f(d, \boldsymbol{\theta}) > f(d_1, \boldsymbol{\theta}) + \Delta\}, \quad (2)$$

where we assume that a beneficial effect is associated with larger values of the response variable. Note that the MED may not exist, as no dose in  $(d_1, d_k]$  may produce an improvement of  $\Delta$  compared with placebo.

#### 3.1 Robust designs for MED estimation

Given a function  $f^0$ , it follows from (2) that the MED (provided it exists) is a solution to

$$\theta_0 + \theta_1 f^0(0, \boldsymbol{\theta}^0) + \Delta = \theta_0 + \theta_1 f^0(\text{MED}, \boldsymbol{\theta}^0). \quad (3)$$

Consequently,  $\text{MED} = b(\boldsymbol{\theta}) = h^0(f^0(0, \boldsymbol{\theta}^0) + \Delta/\theta_1)$ , where  $h^0(x) = \inf\{z | f^0(z) \geq x\}$  denotes the (generalized) inverse of the function  $f^0$  with respect to the variable  $d$ . Standard asymptotic theory for non-linear models (Seber & Wild 1989) yields that the maximum likelihood (ML) estimate

$\widehat{\boldsymbol{\theta}}$  is approximately multivariate normal distributed with mean vector  $\boldsymbol{\theta}$  and covariance matrix  $\frac{\sigma^2}{N} \mathbf{M}^{-1}(\boldsymbol{\theta}, \mathbf{w})$ , where  $\mathbf{M}(\boldsymbol{\theta}, \mathbf{w}) = \sum_{i=1}^k w_i g(d_i, \boldsymbol{\theta}) g(d_i, \boldsymbol{\theta})'$  denotes the information matrix and  $g(d, \boldsymbol{\theta}) = \frac{\partial f(d, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  the gradient of the dose-response model  $f$  with respect to  $\boldsymbol{\theta}$ . It follows from the  $\delta$ -method (van der Vaart 1998) that the MED estimator based on  $\widehat{\boldsymbol{\theta}}$ ,  $\widehat{\text{MED}} = b(\widehat{\boldsymbol{\theta}})$ , is asymptotically normally distributed with mean  $b(\boldsymbol{\theta})$  and variance  $V(\boldsymbol{\theta}, \mathbf{w}) = \frac{\sigma^2}{N} \nabla b(\boldsymbol{\theta})' \mathbf{M}^{-1}(\boldsymbol{\theta}, \mathbf{w}) \nabla b(\boldsymbol{\theta})$ , where  $\nabla b(\boldsymbol{\theta}) = \frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ . Hence, minimizing  $V(\boldsymbol{\theta}, \mathbf{w})$  with respect to  $\mathbf{w} \in \mathbb{S}^k = \{\mathbf{w} \mid \sum_{i=1}^k w_i = 1, \mathbf{w} \geq 0\}$  results in optimal designs that minimize the approximate variance of  $\widehat{\text{MED}}$ . This design criterion has also an appealing decision theoretic justification: The asymptotic normal distribution of  $\widehat{\text{MED}}$  approximates the posterior distribution of the MED in a Bayesian model framework. Hence, minimizing the log-variance of  $\widehat{\text{MED}}$  is equivalent to minimizing the (approximate) Shannon entropy of the posterior distribution of the MED (Chaloner & Verdinelli 1995).

In principle, the above optimization could be done with respect to the number and choice of doses and their corresponding allocation ratios (Dette et al. (2008)), but in practice manufacturing constraints often determine the available doses, as it was the case in the asthma study from Section 2. In the following we thus restrict the optimization to the weights  $\mathbf{w}$  for pre-specified doses  $d_1, \dots, d_k$ .

The true dose-response function  $f$  is unknown and optimal designs are typically not robust with respect to model misspecification (Dette et al. (2008)). In the following we assume a set of  $M$  candidate models  $f_m(d, \boldsymbol{\theta}_m) = \theta_{0m} + \theta_{1m} f_m^0(d, \boldsymbol{\theta}_m^0)$ ,  $m = 1, \dots, M$ , such as those described in Table 1. We “integrate” the design criterion conditional on model  $m$  with respect to the model probabilities  $\alpha_m$ . Hence, using the design criterion  $\sum_{m=1}^M \alpha_m \log(V_m(\boldsymbol{\theta}_m, \mathbf{w}))$  or, equivalently,

$$\Psi(\mathbf{w}) = \prod_{m=1}^M (V_m(\boldsymbol{\theta}_m, \mathbf{w}))^{\alpha_m} \quad (4)$$

leads to designs that are robust with respect to model misspecification, where  $V_m(\boldsymbol{\theta}_m, \mathbf{w})$  denotes the variance of the estimate for the MED in the  $m$ th model ( $m = 1, \dots, M$ ). Note that because of taking logarithms above there is no need to standardize the individual model variances. However, the numerical calculation of robust designs using the criterion (4) requires the knowledge of  $\boldsymbol{\theta}_m$  and  $\alpha_m$ ,  $m = 1, \dots, M$ . In the following sections we describe how the initial best parameter guesses can be updated during an ongoing study such that subsequent stages can be re-designed



based on the updated estimates for  $\theta_m$  and  $\alpha_m, m = 1, \dots, M$ ; see Section 3.3 for a description of the complete procedure in an algorithmic form.

## 3.2 Updating of model parameters and weights

Reliably estimating the parameters  $\theta_1, \dots, \theta_M$  is a challenging problem, particularly in early stages of a study. ML estimates for these parameters are typically highly variable, and may even not exist without imposing bounds on the parameter space. One way of stabilizing estimates is to use a shrinkage approach based on, for example, penalized maximum likelihood or maximum a-posteriori (MAP) estimates. Here, one optimizes the log-likelihood function plus a term which determines the prior plausibility of the parameters (the log prior distribution). The estimate is then a compromise between the information contained in data and the prior distribution. This stabilizes the estimates in early stages due to the shrinkage towards a-priori reasonable values. In later stages the shrinkage effect decreases because the log prior remains constant while the log likelihood receives more weight with increasing sample sizes. If a completely flat prior distribution is used, standard ML and MAP estimation coincide, so that using non-uniform priors is desirable. We discuss the choice of non-uniform priors in more detail further below.

Apart from stable parameter estimates  $\theta_1, \dots, \theta_M$  for the dose-response models, one needs to update the model probabilities  $\alpha_1, \dots, \alpha_M$  at an interim analysis. We propose using a probability distribution over the different dose-response models and evaluating the posterior probabilities for each model after having observed the data; see, for example, Kass & Raftery (1995) for a detailed description of posterior probabilities and Bayes factors. These posterior model probabilities can then be used in the design criterion (4). A computationally efficient approach to approximate the posterior model probabilities is the Bayesian information criterion (BIC). However, previous simulation studies in the context of dose-finding studies showed that the BIC approximation frequently favors too simplistic models for realistic variances and sample sizes (Bornkamp 2006). Other approximate methods, such as fractional Bayes factors, (O'Hagan 1995) or intrinsic Bayes factors (Berger & Pericchi 1996) either depend on arbitrary tuning parameter values or are computationally prohibitive. Thus, for each model we will use the exact posterior probabilities

resulting from the prior distributions assumed for the MAP estimation. In our case, these probabilities can be calculated using efficient numerical quadrature without the need to resort to computationally expensive Markov Chain Monte Carlo techniques. In the remainder of this section we provide details on the prior elicitation and the calculation of posterior probabilities.

### 3.2.1 Selection of prior distributions for $\theta_m$

We utilize the factorization in (1) to derive a prior distribution for  $(\theta_{0m}, \theta_{1m}, \theta_m^0, \sigma^2)$ . If  $\theta_m^0$  were known, the non-linear models would reduce to a linear model. It is therefore reasonable to use for a given  $\theta_m^0$  the conditionally conjugate normal-inverse gamma (NIG) distribution

$$p(\theta_m^*, \sigma^2 | \theta_m^0) \propto (\sigma^2)^{-(d+4)/2} \exp[-\{(\theta_m^* - \boldsymbol{\mu})' \mathbf{V}^{-1} (\theta_m^* - \boldsymbol{\mu}) + a\} / (2\sigma^2)]$$

for  $(\theta_m^*, \sigma^2)$  (O'Hagan & Forster 2004), where  $a, d > 0$ ,  $\boldsymbol{\mu} \in \mathbb{R}^2$  and  $\mathbf{V} \in \mathbb{R}^{2 \times 2}$  denotes a positive definite matrix. The NIG distribution marginally induces a bivariate t-distribution for  $\theta_m^*$  with  $d$  degrees of freedom, finite mean  $\boldsymbol{\mu}$  and covariance matrix  $a/(d-2)\mathbf{V}$ , provided that  $d > 2$ . The marginal prior distribution for  $\sigma^2$  is given by an IG distribution with mode  $a/(d+2)$ , mean  $a/(d-2)$  and variance  $2a^2/\{(d-2)^2(d-4)\}$ . It has a finite mean when  $d > 2$  and a finite variance when  $d > 4$ .

To set up the NIG distribution for  $(\theta_m^*, \sigma^2)$  one can employ available information about the placebo effect, the maximum treatment effect and the standard deviation. For example, one can choose the marginal bivariate t-distribution for  $\theta_m^*$  (conditional on  $\theta_m^0$ ) such that the desired mean and covariance are achieved for the placebo effect and the maximum effect of the underlying dose-response model. When the linear parameters  $\theta_m^*$  cannot be interpreted as placebo and maximum effect, one can use a suitable transformation to achieve the desired moments. Then, one can adjust  $a$  and  $d$  so that the marginal distribution of  $\sigma^2$  achieves the desired mode. An attractive choice is to use  $d = 4$ , leading to a prior with infinite variance for  $\sigma^2$  and a heavy tailed marginal prior for  $\theta_m^*$ .

For the non-linear parameters  $\theta_m^0$ , we propose selecting suitable bounds for the parameters and then eliciting a bounded prior distribution. This is typically not difficult, as the interpretation

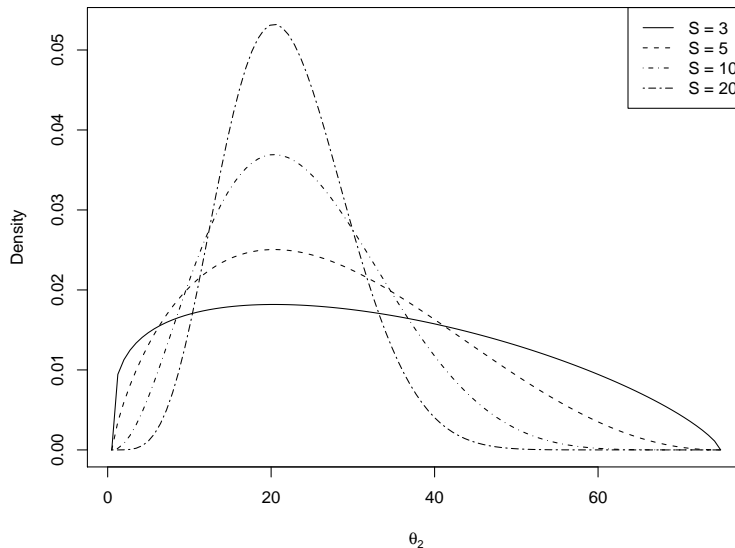


Figure 2: *Beta priors on  $[0.5, 75]$  with mode 20 and different  $S$  values.*

of the non-linear parameters is straightforward, and excessively large parameter values usually correspond to a-priori unlikely model shapes. We propose using a scaled beta distribution  $B(\alpha, \beta)$  with mode equal to the initial parameter guesses. The curvature of the prior determines the amount of shrinkage that one is willing to employ for the MAP estimates. In the simulations we used the sum  $S = \alpha + \beta$  as a measure of curvature with  $S > 2$  to ensure unimodality of the beta distribution. Note that already relatively small values of  $S$ , such as  $S = 10$  or  $S = 20$ , lead to strong shrinkage effects, see Figure 2 for an illustration of the  $\theta_2$  parameter in the  $\text{E}_{\max_1}$  model, where the initial parameter guess is 20. For dose-response models with more than one non-linear parameter, we repeat this procedure for all parameters and assume independence among them.

For selecting prior model probabilities, it is convenient to use a uniform distribution across the models unless some models are deemed a-priori more plausible than others.

### 3.2.2 Calculation of posterior probabilities

Let  $\mathbf{y}$  denote the data available at an interim analysis and  $p(\mathbf{y}|\boldsymbol{\theta}_m, m)$  the likelihood under model  $m$  with corresponding prior distribution  $p(\boldsymbol{\theta}_m|m)$  and prior model probability  $p(m)$ . Then the

marginal likelihood is given by

$$\begin{aligned} p(m|\mathbf{y}) &\propto p(m) \int p(\mathbf{y}|\boldsymbol{\theta}_m, \sigma^2, m) p(\boldsymbol{\theta}_m, \sigma^2|m) d(\boldsymbol{\theta}_m, \sigma^2) \\ &= p(m) \int \int p(\mathbf{y}|\boldsymbol{\theta}_m, \sigma^2, m) p(\boldsymbol{\theta}_m^*, \sigma^2|\boldsymbol{\theta}_m^0, m) d(\boldsymbol{\theta}_m^*, \sigma^2) p(\boldsymbol{\theta}_m^0|m) d\boldsymbol{\theta}_m^0. \end{aligned}$$

The inner integral in the last equation is the product of a likelihood and a conjugate prior distribution. One can hence reduce the integration to

$$p(m|\mathbf{y}) \propto p(m) \int p(\mathbf{y}|\boldsymbol{\theta}_m^0, m) p(\boldsymbol{\theta}_m^0|m) d\boldsymbol{\theta}_m^0, \quad (5)$$

where  $p(\mathbf{y}|\boldsymbol{\theta}_m^0, m)$  now denotes the integrated likelihood. In our applications, the integral (5) is one- or two-dimensional over a bounded region and hence straightforward to calculate numerically. This allows us to calculate the marginal likelihoods efficiently, without resorting to Markov Chain Monte Carlo calculations, see Section 3.4 for details. The posterior model probabilities can be obtained by normalizing the marginal likelihoods in (5) by their sum.

We use the maximum  $\tilde{\boldsymbol{\theta}}_m^0$  of the marginal posterior  $p(\mathbf{y}|\boldsymbol{\theta}_m^0, m) p(\boldsymbol{\theta}_m^0|m)$  as an estimate of  $\boldsymbol{\theta}_m^0$ . Conditional on this value, we use the maximum  $\tilde{\boldsymbol{\theta}}_m^*$  of  $p(\boldsymbol{\theta}_m^*|\tilde{\boldsymbol{\theta}}_m^0, \mathbf{y}, m)$  as an estimate for  $\boldsymbol{\theta}_m^*$ . Therefore, the overall estimate of the parameter  $\boldsymbol{\theta}_m$  is given by  $\tilde{\boldsymbol{\theta}}_m = (\tilde{\boldsymbol{\theta}}_m^*, \tilde{\boldsymbol{\theta}}_m^0)$ . This is a slight variation of the MAP approach described above, but reduces further the computational effort as it re-uses the calculations from the integration to obtain the marginal likelihoods.

### 3.3 Main algorithm

We now summarize the complete response-adaptive dose-finding design in algorithmic form.

*Before Trial Start*

1. Select a starting design using either a balanced allocation across the available doses or an unbalanced allocation based on optimal design considerations.
2. Select candidate dose-response models  $f_m(d, \boldsymbol{\theta}_m)$ .

3. Conditional on  $\theta_m^0$ , calculate a NIG prior distribution for  $(\theta_m^*, \sigma^2)$  based on “best guesses” for the placebo effect, the maximum treatment effect and  $\sigma^2$  (together with suitable variability assumptions for both parameters).
4. Choose “best guesses” for the non-linear parameters  $\theta_m^0$  and select the parameter  $S$ .
5. Choose prior model probabilities  $p(m)$  for the different dose-response functions.

*At Interim Analysis*

1. Calculate posterior model probabilities

$$p(m|\mathbf{y}) \propto p(m) \int p(\mathbf{y}|\theta_m, m)p(\theta_m|m)d\theta_m. \quad (6)$$

Exploiting the conjugacy properties of the NIG distribution, this reduces to one- or two-dimensional integrals, see Section 3.4 for details.

2. For each dose-response model, estimate  $\theta_m$  by using the maximum of  $p(\mathbf{y}|\theta_m^0, m)p(\theta_m^0|m)$ , where the abscissas calculated in step 1 can be re-used. Conditional on this value, use the maximum of  $p(\theta_m^*|\tilde{\theta}_m^0, \mathbf{y}, m)$  as an estimate for  $\theta_m^*$  to obtain  $\tilde{\theta}_m$
3. Plug the obtained parameter estimates  $\tilde{\theta}_m$  into (4) and set  $\alpha_m = p(m|\mathbf{y})$ . Then, minimize with respect to  $\mathbf{w} \in \mathcal{S}$ , where  $\mathcal{S} = \{\mathbf{w} \in \mathbb{S}^k | \mathbf{w} = (\mathbf{n}_{old} + N_{next}\mathbf{w}_{next})(N_{old} + N_{next})^{-1}, \mathbf{w}_{next} \in \mathbb{S}^k\}$ . Here,  $\mathbf{n}_{old}$  denotes the vector of sample sizes per dose and  $N_{old}$  the total sample size until the current interim analysis. Further,  $N_{next}$  denotes the sample size and  $\mathbf{w}_{next} \in \mathbb{S}^k$  the design weights for the next cohort of patients. We therefore optimize the design for the next stage taking into account the patient allocation until the current interim analysis, see Section 3.4 for computational details.
4. Allocate the next cohort of patients according to  $\mathbf{w}_{next}$  by applying an appropriate rounding technique, such as described in Pukelsheim (1993, ch. 12).

Note that the Bayesian approach is used here for design adaptation purposes. The final analysis may or may not be done using a fully Bayesian approach. The development of the Bayesian design

methodology above is motivated by the MCP-Mod methodology described in (Bretz, Pinheiro & Branson 2005) to address model uncertainty. This method requires prior estimates for the placebo effect, the maximum treatment effect, and  $\sigma^2$  at the design stage and “best guesses” of the non-linear parameters  $\boldsymbol{\theta}_m^0$  for the analysis. The additional information needed to set up the above adaptive design procedure is hence minimal. Obviously, any other strategy that allows one to use a set of candidate dose-response models might also be used for the final analysis.

### 3.4 Technical details

In this section we provide further details of the algorithm presented above.

For the calculation of the one- and two-dimensional integrals in (5) we used quasi-uniformly distributed point sets based on good lattice points  $\mathbf{u}_1, \dots, \mathbf{u}_n$ , where  $\mathbf{u}_i \in [0, 1]^d$  and  $d$  is the dimension of  $\boldsymbol{\theta}_m^0$ ; see Fang & Wang (1994) for details on the construction of such integration grids. Let  $\pi(\boldsymbol{\theta}_m^0|y) = p(\mathbf{y}|\boldsymbol{\theta}_m^0, m)p(\boldsymbol{\theta}_m^0|m)$  denote the integrand from (5) and let  $\mathbf{b}_l$  and  $\mathbf{b}_u$  denote the vector of lower and upper bounds for  $\boldsymbol{\theta}_m^0$ . One first transforms the good lattice points to obtain  $\mathbf{u}_i^* = \mathbf{u}_i(\mathbf{b}_u - \mathbf{b}_l) + \mathbf{b}_l$  for  $i = 1, \dots, n$ , and then approximates the integral (5) by  $\prod_{j=1}^d (b_{uj} - b_{lj}) \sum_{i=1}^n \pi(\mathbf{u}_i^*|y)/n$ . This approach also allows one to calculate the approximate maximum in the subsequent optimization step by using the grid point  $\mathbf{u}_i^*$  corresponding to  $\max_{i \in \{1, \dots, n\}} \pi(\mathbf{u}_i^*|y)$ . We found that using grids of size 100 in the one-dimensional case and 1597 in the two-dimensional case provide reliable and computationally efficient results (for integration and optimization).

The optimization in (4) is a constrained optimization problem because the weights  $\mathbf{w}_{next}$  lie in the  $(k - 1)$ -dimensional probability simplex. A simple but efficient approach to perform the optimization is to use a mapping  $\mathbb{R}^{k-1} \mapsto \mathbb{S}^k$  and then to employ a standard unconstrained optimizer, as described in Atkinson, Donev & Tobias (2007, p. 131). To account for the already allocated patients until an interim analysis, one can optimize  $\Psi \left( \frac{n_{old} + N_{next} \mathbf{w}_{next}}{N_{old} + N_{next}} \right)$  with respect to  $\mathbf{w}_{next} \in \mathbb{S}^k$ . Due to potential multiple optima in the design surface, one cannot be sure, whether indeed an optimal design has been found by the optimizer. We thus propose using lower

bounds of the resulting relative efficiencies based on the underlying geometry of the optimization problem.

To be precise, suppose that the vector  $\mathbf{w}^*$  has been found by the optimizer. The following result gives a lower bound on  $r(\mathbf{w}^*) = \frac{\Psi(\mathbf{w}_{opt})}{\Psi(\mathbf{w}^*)} \in [0, 1]$ , where  $\mathbf{w}_{opt}$  is the (unknown) true optimal design at the end of the next stage, accounting for the patients allocated until the current interim analysis. A proof of the result is given in the Appendix.

**Theorem 1.** A design  $\mathbf{w}$  with  $c_m(\boldsymbol{\theta}_m) = \nabla b_m(\boldsymbol{\theta}_m) \in \text{Range}(\mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}))$ ,  $m = 1, \dots, M$ , minimizes  $\Psi(\mathbf{w})$  with respect to  $\mathbf{w}_{next}$ , where  $\mathbf{w} = \frac{N_{old} + N_{next} \mathbf{w}_{next}}{N_{old} + N_{next}}$ , if and only if there exist generalized inverses  $\mathbf{G}_1, \dots, \mathbf{G}_m$  of  $\mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w})$ , such that the inequality

$$h(d, \mathbf{w}) = \frac{\sum_{m=1}^M \alpha_m \frac{(g_m^T(d, \boldsymbol{\theta}_m) \mathbf{G}_m c_m(\boldsymbol{\theta}_m))^2}{c_m^T(\boldsymbol{\theta}_m) \mathbf{G}_m c_m(\boldsymbol{\theta}_m)}}{\sum_{m=1}^M \alpha_m \frac{c_m(\boldsymbol{\theta}_m) \mathbf{G}_m^T \mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}_{next}) \mathbf{G}_m c_m(\boldsymbol{\theta}_m)}{c_m^T(\boldsymbol{\theta}_m) \mathbf{G}_m c_m(\boldsymbol{\theta}_m)}} \leq 1$$

is satisfied for all  $d \in \{d_1, \dots, d_k\}$ . Moreover, the efficiency of any design  $\mathbf{w}$  can be bounded from below by

$$r(\mathbf{w}) \geq \frac{1}{k^*(\mathbf{w}, \gamma)} \geq \frac{1}{h^*(\mathbf{w})}, \quad (7)$$

where  $\gamma = \frac{N_{old}}{N_{old} + N_{next}}$ ,  $h^*(\mathbf{w}) = \min_{\mathbf{G}_1, \dots, \mathbf{G}_m} \max_{d \in \{d_1, \dots, d_k\}} h(d, \mathbf{w})$ ,

$$k^*(\mathbf{w}, \gamma) = \left[ 1 + (1 - \gamma) \min_{\mathbf{G}_1, \dots, \mathbf{G}_m} \max_{\mathbf{v} \in \mathbb{S}^k} \sum_{m=1}^M \alpha_m \frac{c_m^T(\boldsymbol{\theta}_m) \mathbf{G}_m^T (\mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{v}) - \mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}_{next})) \mathbf{G}_m c_m(\boldsymbol{\theta}_m)}{c_m^T(\boldsymbol{\theta}_m) \mathbf{G}_m c_m(\boldsymbol{\theta}_m)} \right]^{-1},$$

and the minimum is taken over all generalized inverses  $\mathbf{G}_1, \dots, \mathbf{G}_m$  of the matrices  $\mathbf{M}_1(\boldsymbol{\theta}_1, \mathbf{w})$ ,  $\dots$ ,  $\mathbf{M}_M(\boldsymbol{\theta}_M, \mathbf{w})$ .

When the matrices  $\mathbf{M}_1(\boldsymbol{\theta}_1, \mathbf{w})$ ,  $\dots$ ,  $\mathbf{M}_M(\boldsymbol{\theta}_M, \mathbf{w})$  are invertible,  $h^*(\mathbf{w})$  is just the maximum of  $h(d, \mathbf{w})$  over the  $k$  doses and straightforward to calculate (and so is the lower bound on  $r(\mathbf{w})$ ). This lower bound is useful in several respects: We do not need to know the actual optimal design  $\mathbf{w}_{opt}$  in order to calculate the lower bound. If the lower bound for our calculated design  $\mathbf{w}^*$  is equal to 1, we know that  $\mathbf{w}^*$  is the optimal design. Otherwise, we have a conservative estimate on how much percent off one would be when using  $\mathbf{w}$ . The bound based on  $k^*(\mathbf{w}, \gamma)$  is sharper, however harder to implement.

If one does not use a fully Bayesian approach for the final analysis, one typically has to fit non-linear regression models to the data. When there are only few doses available, as it is often the case in drug development practice, calculating the ML estimate may be difficult. One way to simplify the problem is by exploiting the fact that  $\theta_0$  and  $\theta_1$  enter the model function linearly in (1). We thus apply the nonlinear optimization only on the non-linear parameters  $\boldsymbol{\theta}_m^0$ , similar in spirit to Golub & Pereyra (2003). Using the Frisch-Waugh-Lovell theorem (Baltagi 2008, ch. 7) we can recalculate the residual sum of squares efficiently, without the need to solve the full least squares problems in each iteration of the non-linear optimization (this effect becomes even more important when there are additional linear covariates in the model equation, such as gender, baseline values, etc.). In addition, we impose bounds on the non-linear parameters  $\boldsymbol{\theta}_m^0$  to guarantee the existence of the least squares estimate (Seber & Wild 1989, chapter 12). As mentioned in Section 3.2.1, such bounds are not a severe restriction in practice and ensure that the optimization problem is well-posed.

## 4 Asthma study re-visited

In this section we re-visit the asthma case study from Section 2 and address the four open design questions using the proposed methodology from Section 3. To this end, we investigated in an extensive simulation study the operating characteristics for different design options and parameter configurations.

### 4.1 Design of simulation study

We generated normally distributed observations according to the dose-response models given in Table 1 with  $\sigma = 350$  mL. To investigate the robustness of the proposed methods, we also simulated from a linear model (with baseline 100 mL and maximum effect 300 mL) that was not included in the candidate model set. The total sample size was fixed at 300 (constraint imposed by the clinical team). To evaluate the benefit of including additional doses we compared two



design options, one with the four active doses 2.5, 10, 20, 50 (plus placebo) and another one with the seven active doses 0.5, 1, 2.5, 5, 10, 20 and 50 (plus placebo). In addition, we evaluated the benefit of additional interim analyses by varying their number from 0 (= no interim analysis) to 9, where the interim looks were chosen equally spaced in time. In all cases, we assumed a balanced first stage design. The designs from the second stage onwards were determined using the observed data according to the algorithm from Section 3.3. When the MED did not exist for certain models at an interim analysis, they were removed from the model set and the model probabilities were accordingly re-weighted. When the MED did not exist for any model, balanced allocations were used for the next cohort of patients.

For the final analysis we employed the MCP-Mod procedure from Bretz, Pinheiro & Branson (2005). A potential dose-response signal was assessed using model-based multiple contrast tests based on the candidate model set from Table 1. Subsequently, if there were significant models, the dose-response model with lowest Akaike Information Criterion (AIC) among the significant models was chosen to estimate the MED.

The methodology from Section 3 was applied with uniform prior probabilities for the different models. We further assumed a-priori distributions with mean 100 and variance 100000 for the placebo effect and mean 300 and variance 100000 for the maximum treatment effect, which were then transformed into the linear parameters for all dose-response models. The mode of the marginal distribution for  $\sigma^2$  was chosen as  $350^2$  with  $d = 4$ , resulting in an infinite variance. For the non-linear parameters we assumed beta distributions (or products thereof) with mode equal to the values specified in Table 1 and  $S = 3$ . The parameter bounds were chosen to ensure that all reasonable dose-response shapes remained included within the bounds. That is, we chose  $\theta_2 \in [0.05, 75]$  for the Emax models,  $(\theta_2, \theta_3) \in [0.5, 4] \times [0.5, 4]$  for the beta models and  $(\theta_2, \theta_3) \in [0.05, 75] \times [0.5, 25]$  for the the logistic model. For each scenario we used 5000 simulation runs.

## 4.2 Simulation results

For the chosen standard deviation of  $\sigma = 350$  mL the power of the MCP-Mod procedure to detect a dose-response signal was almost always close to 1. Thus, the MCP-Mod procedure was essentially reduced to choosing the non-linear model with lowest AIC value under the constraint that only models with significant contrast test statistics were included in the model selection step. Simulations with  $\sigma$  values larger than 350 mL indicated that the power quickly dropped to lower levels (results not shown here), although the estimation results remained qualitatively similar to the ones shown below.

In Figure 3 we display the mean absolute estimation error for the MED estimates against the number of interim analyses. In all scenarios one observes a benefit from adapting, while most of the improvement is already achieved after 1, 2 or 4 interim analyses. The largest relative improvement (comparing no adaptations vs 9 adaptations) can be observed for the Logistic<sub>1</sub> and the Beta model, particularly in the case of 7 active doses. The worst relative improvement can be observed for the Logistic<sub>2</sub> model, which overall has the largest absolute estimation error. This is not surprising, because even when adapting one cannot achieve a good design for this model, as there are no doses available for administration in the interval (20, 50) containing the MED, see also Figure 1. It is remarkable to see that adaptation also works if the true model is a linear model, although it is not included in the candidate model set. It seems that other models in the candidate set are able to capture the shape of a linear model reasonably well.

The comparison between 4 and 7 active doses is not entirely clear. If no interim analyses are performed, it seems that the design with a balanced allocation across the 4 active doses is slightly better than the design with a balanced allocation across all 7 active doses. If one decides to adapt, however, it seems beneficial in some cases to have more doses available, particularly if many interim analyses are performed, while in other cases 4 active doses are sufficient.

To illustrate how adaptation changes the allocation of patients to the different doses, we display in Figure 4 the average patient allocations for the Emax<sub>2</sub> model after 1, 2, 4, and 9 interim analyses and with 7 available active doses. The adaptive design tends to allocate more patients

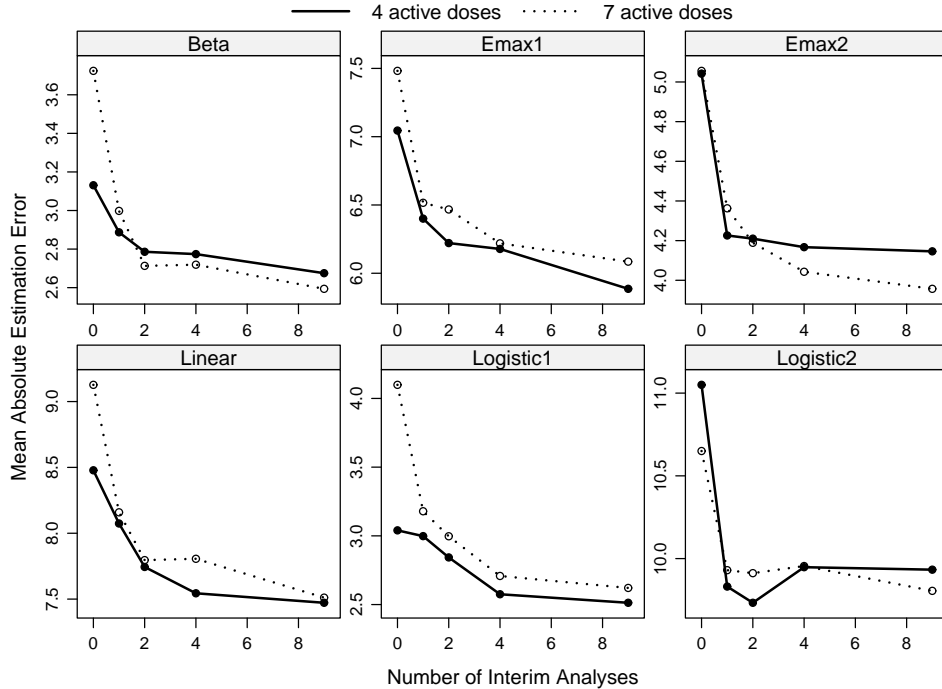


Figure 3: Mean absolute estimation error for MED estimation.

both on placebo and nearby the actual MED. This is intuitively plausible, as the MED estimate depends on the precision of the estimated placebo effect as well as of the estimated function  $f(\cdot)$  around the true MED. It also follows from Figure 4 that for a large number of interim analyses the overall allocation is close to the one under a locally optimal design for the  $E_{\max_2}$  model, with the variability in the allocations due to the uncertainty both in estimating the correct model and the model parameters at the interim analysis. Similar conclusions also hold for other models than the  $E_{\max_2}$  model (not reported here).

We now investigate to which extent the precision gain observed in Figure 3 translates into sample size savings when performing an adaptive design. In other words, how many additional patients are required for a non-adaptive, balanced design to achieve a similar estimation error as with an adaptive design using 300 patients. We again considered the  $E_{\max_2}$  model and iterated the total sample size until the mean absolute estimation error was approximately 4 (which is the mean absolute estimation error obtained after 9 interim analyses, as seen in Figure 3). For both design options with 4 and 7 active doses this was achieved after roughly 500 patients. Thus,

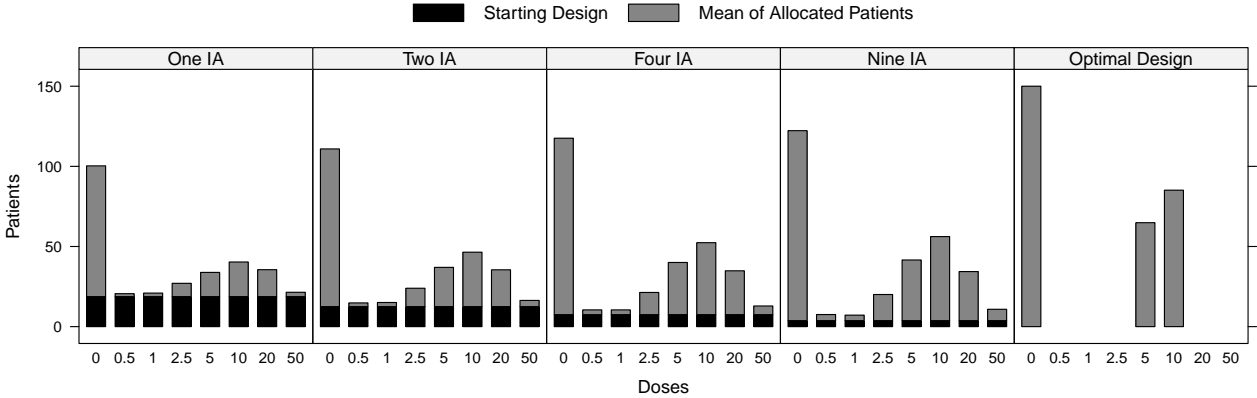


Figure 4: Average patient allocation after 1, 2, 4, and 9 interim analysis (IA) under the  $E_{max_2}$  model. Last panel: locally MED-optimal design for the  $E_{max_2}$  model with true MED = 7.7.

using a non-adaptive, balanced design one would need 200 additional patients to achieve a similar precision in MED estimation as compared to an adaptive design using 300 patients.

The adaptive design benefits observed so far depend on several input parameters, such as the starting design for the first stage. One may argue that starting with a bad design that allocates patients at the “wrong” doses may be improved by adapting at one or more interim looks. On the other hand, starting with a good design may lead to adaptations following random noise at the interim analyses. To illustrate this effect, we report the results for the simulations under the  $\text{Logistic}_1$  model (similar results were also obtained for other models and scenarios, but are not reported here). We used four different starting designs. We used  $\mathbf{w} = (0.35, 0.03, 0.22, 0.35, 0.05)$  and  $\mathbf{w} = (0.35, 0.02, 0.02, 0.02, 0.02, 0.20, 0.30, 0.07)$  as good starting designs with 4 and 7 active doses, respectively. These designs work well because they allocate patients on placebo and around the MED, while keeping some mass on the remaining doses. In addition, we used  $\mathbf{w} = (0.1, 0.3, 0.05, 0.05, 0.5)$  and  $\mathbf{w} = (0.1, 0.2, 0.22, 0.02, 0.02, 0.02, 0.02, 0.4)'$  as bad starting designs, as they have relatively few patients on placebo and around the MED. It follows from Figure 5 that substantial improvements are possible when using bad starting designs. On the other hand, for good starting designs no benefit is achieved by adapting and the performance may even deteriorate, because the possibility of adapting may lead one to deviate from the already good starting design. In practice, one does not know whether an employed design is good or bad, but

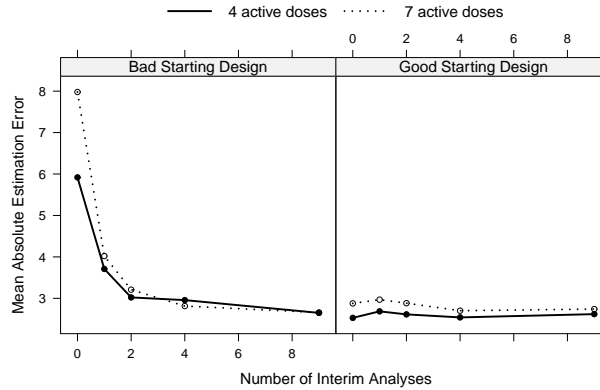


Figure 5: *Mean absolute estimation error for MED estimation, under the the  $Logistic_1$  model for good and bad starting designs.*

one should keep in mind the possibility that adaptive designs will not always improve upon the initial design.

To further investigate the robustness of the proposed methods we repeated the simulation study from Figure 3 by increasing the standard deviation to 450 mL and 700 mL. The overall results remain similar, but with increased absolute estimation errors. However, the relative benefit of 9 interim analyses vs. no adaptation decreases slightly. Due to the larger noise, one obtains less reliable information at an interim analysis and one may end up with a worse design for the next stage. We also investigated the effect of prior misspecification. For this purpose we misspecified the prior means or prior modes by adding or subtracting 20% of the true value, but leaving the variability (variance of baseline and maximum effect and the value  $S$  for the beta distribution) unchanged as in the original simulations. The results are largely identical to those presented in Figure 3, indicating that the proposed methods are robust under moderate prior misspecifications.

### 4.3 Conclusion for asthma study

Many more simulations than presented above were conducted at the planning stage of the asthma study to address the four questions stated in Section 2. Regarding question (A) the team felt

that the potential benefits of conducting an adaptive design (more precise MED estimation) outweighed the additional logistical requirements, especially in view of the perceived sample size gain of 100 - 200 patients when compared to a fixed-sample study designed to achieve a similar precision. For question (B) it was decided to have one interim analysis: Based on Figure 3 and other simulation results, the potential further reduction of the mean absolute estimation error with two or more interim analyses was perceived as too small to justify the additional logistical complexity.

For similar reasons, the team decided against having all seven active doses from the beginning on (question (C)). Instead, it was decided to allocate 150 patients equally across the four active doses 2.5, 10, 20, 50 (plus placebo) in the first stage. Once the interim results are available and analyzed with the methods from Section 3, however, patients could be allocated to all seven active doses (or a subset thereof) in the second stage. For practical reasons, the clinical team decided to incorporate constraints on the minimum number of patients allocated per dose in the second stage: If the algorithm would allocate less than 5% of the patients on a certain dose, that dose would be dropped altogether and the corresponding patients re-allocated to the remaining doses.

## 5 Discussion

Motivated by a dose finding study in moderate persistent asthma, we described a response-adaptive approach that addresses common challenges encountered in dose-finding studies: multiple objectives, model uncertainty, and large variability. When planning an adaptive dose-finding design it is important to realize that it may not always be better than a non-adaptive design. It is necessary to employ a factored view as many parameters may impact the performance of a study design. Often, an unbalanced fixed-sample design derived from optimal design theory might already provide benefit over a balanced fixed-sample design and adaptation may not bring further advantages, particularly if the variability is large (which is common in practice). Thus, adaptive designs are promising in situations where the initial design is not good and/or interim

data have low variability. In practice one never knows how good the initial design will be, before trial start, and adaptive designs may guard against bad initial designs. However, the benefits of adaptive dose-finding designs have to be balanced against the increased logistical requirements to implement processes for repeated data collection, cleaning and analyses, to maintain trial integrity and validity, and to overcome potential challenges in drug manufacturing and supply.

In this paper we focused on designs based on the compound optimality criterion (4) to address model uncertainty and to minimize the variance of  $\widehat{\text{MED}}$ . The criterion depends on the parameters of the different dose-response models as well as on the model probabilities and we used a Bayesian approach to continuously update parameter values and model probabilities based on the information accrued in the trial. The approach was implemented based on optimization and numerical quadrature, so that computationally intensive Markov Chain Monte Carlo techniques could be avoided. Computational efficiency is of extreme importance as the frequentist operating characteristics of any adaptive design methodology needs to be evaluated in extensive simulations under multiple scenarios.

The proposed method can be extended immediately if alternative optimality criteria (such as  $\text{ED}_p$ - or D-optimal designs, see Dette, Kiss, Bevanda & Bretz (2010)) or mixtures thereof are of interest. Alternatively, optimal discrimination designs could be applied that allow one to differentiate among several candidate nonlinear regression models (Atkinson & Fedorov (1975), Dette & Titoff (2009)). It would be interesting to address multiple objectives by considering different optimality criteria at different stages, such as using a model discrimination design in earlier stages, and MED-optimal design in later stages. This will be investigated in future research, but see Dragalin et al. (2010) for initial results.

The R code used for the simulations is largely available with the `DoseFinding` package in R Bornkamp, Pinheiro & Bretz (2010). Currently the package includes functions for calculating and updating optimal designs and an implementation of the MCP-Mod procedure. The Bayesian updating step is available from the authors upon request.

**Acknowledgements** The authors would like to thank Martina Stein, who typed parts of this

manuscript with considerable technical expertise. The work of Holger Dette has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research Foundation (DFG).

## Appendix: Proof of Theorem 1

**Proof of Theorem 1.** Obviously the first part of the Theorem follows from the lower bound (7) on the efficiency. For a proof of (7) let  $\gamma = N_{old}/(N_{old} + N_{next}) \in (0, 1)$  and note that the total information of the experiment in the  $m$ th model is given by

$$\mathbf{M}_m(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{w}_{next}) = \gamma \mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}_{old}) + (1 - \gamma) \mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}_{next}) \quad (8)$$

where we collect in the vector  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M)$  the parameters of the different models. Define a block diagonal matrix by

$$\overline{\mathbf{M}}(\boldsymbol{\theta}_m, \mathbf{w}_{old}, \mathbf{w}_{next}) = \text{diag}(\mathbf{M}_1(\boldsymbol{\theta}_1, \mathbf{w}_{old}, \mathbf{w}_{next}), \dots, \mathbf{M}_M(\boldsymbol{\theta}_M, \mathbf{w}_{old}, \mathbf{w}_{next})) \quad (9)$$

(all other entries in this matrix are 0) and similarly

$$\mathbf{K} = \text{diag}(c_1(\boldsymbol{\theta}_1), \dots, c_M(\boldsymbol{\theta}_M)),$$

where the vector  $c_m(\boldsymbol{\theta})$  is given by  $c_m(\boldsymbol{\theta}) = \nabla b_m(\boldsymbol{\theta}), m = 1, \dots, M$ . For a design  $\mathbf{w}_{next}$ , such that  $c_m(\boldsymbol{\theta}) \in \text{Range}(\mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}_{old}, \mathbf{w}_{next}))$  ( $m = 1, \dots, M$ ) we consider the information matrix

$$\begin{aligned} C_K(\overline{\mathbf{M}}(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{w}_{next})) &= (\mathbf{K}^T \overline{\mathbf{M}}^{-1}(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{w}_{next}) \mathbf{K})^{-1} \\ &= \text{diag}((c_1^T(\boldsymbol{\theta}_1) \mathbf{M}_1^{-1}(\boldsymbol{\theta}_1, \mathbf{w}_{old}, \mathbf{w}_{next}) c_1(\boldsymbol{\theta}_1))^{-1}, \dots, \\ &\quad (c_M^T(\boldsymbol{\theta}_M) \mathbf{M}_M^{-1}(\boldsymbol{\theta}_M, \mathbf{w}_{old}, \mathbf{w}_{next}) c_M(\boldsymbol{\theta}_M))^{-1}). \end{aligned}$$

Note that the optimal design maximizes

$$\begin{aligned} \Psi^{-1}(\mathbf{w}_{next}) &= \frac{N_{old} + N_{next}}{\sigma^2} \cdot \Phi_\alpha(C_K(\overline{\mathbf{M}}(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{w}_{next}))) \\ &= \frac{N_{old} + N_{next}}{\sigma^2} \prod_{m=1}^M (c_m^T(\boldsymbol{\theta}_m) \mathbf{M}_m^{-1}(\boldsymbol{\theta}_m, \mathbf{w}_{old}, \mathbf{w}_{next}) c_m(\boldsymbol{\theta}_m))^{-\alpha_m}, \end{aligned}$$



where the last identity defines the criterion  $\Phi_\alpha$  and we have used the notation  $\Phi_\alpha(\text{diag}(\lambda_1, \dots, \lambda_M)) = \prod_{m=1}^M \lambda_m^{\alpha_m}$ . Now according to Theorem 1 in Dette (1996) a lower bound for the efficiency of the design  $\mathbf{w}_{next}$

$$r(\mathbf{w}) = \frac{\psi^{-1}(\mathbf{w})}{\psi^{-1}(\mathbf{w}_{opt})} = \frac{\Phi_\alpha(C_k(\overline{\mathbf{M}}(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{w}_{next})))}{\max_{\mathbf{v} \in \mathbb{S}^k} \Phi_\alpha(C_K(\overline{\mathbf{M}}(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{v})))}$$

is obtained as

$$e = \left[ \min_{\mathbf{G}} \max_{A \in \mathcal{A}} \text{tr} \left\{ \mathbf{G} \mathbf{K} C_K(\overline{\mathbf{M}}(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{w}_{next})) \mathbf{E} C_K(\overline{\mathbf{M}}(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{w}_{next})) \mathbf{K}^T \mathbf{G}^T A \right\} \right]^{-1}$$

where the minimum is taken over the set of all generalized inverses of the matrix  $\overline{\mathbf{M}}(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{w}_{next})$  and the set  $\mathcal{A}$  is defined by

$$\mathcal{A} = \{ \overline{\mathbf{M}}(\boldsymbol{\theta}, \mathbf{w}_{old}, \mathbf{v}) \mid \mathbf{v} \in \mathbb{S}^k \}$$

and the matrix  $\mathbf{E}$  is given by

$$\mathbf{E} = \text{diag}(\alpha_1 c_1^T(\boldsymbol{\theta}) \mathbf{M}_1^{-1}(\boldsymbol{\theta}_1, \mathbf{w}_{old}, \mathbf{v}) c_1(\boldsymbol{\theta}), \dots, \alpha_M c_M^T(\boldsymbol{\theta}) \mathbf{M}_M^{-1}(\boldsymbol{\theta}_M, \mathbf{w}_{old}, \mathbf{v}) c_M(\boldsymbol{\theta})).$$

Therefore, observing the identity

$$\mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}_{old}, \mathbf{v}) = \mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}_{old}, \mathbf{w}_{next}) + (1 - \gamma)(\mathbf{M}(\boldsymbol{\theta}_m, \mathbf{v}) - \mathbf{M}(\boldsymbol{\theta}, \mathbf{w}_{next})),$$

we obtain

$$\begin{aligned} e &= \left[ 1 + (1 - \gamma) \min_{\mathbf{G}_m} \max_{\mathbf{v} \in \mathbb{S}^k} \sum_{m=1}^M \alpha_m \frac{c_m^T(\boldsymbol{\theta}_m) \mathbf{G}_m^T (\mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{v}) - \mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}_{next})) \mathbf{G}_m c_m(\boldsymbol{\theta})}{c_m^T(\boldsymbol{\theta}_m) \mathbf{G}_m c_m(\boldsymbol{\theta}_m)} \right]^{-1} \\ &\geq \left[ \min_{\mathbf{G}_1, \dots, \mathbf{G}_m} \max_{d \in \{d_1, \dots, d_k\}} \frac{\sum_{m=1}^M \alpha_m \frac{(g_m^T(d, \boldsymbol{\theta}_m) \mathbf{G}_m c_m(\boldsymbol{\theta}_m))^2}{c_m^T(\boldsymbol{\theta}_m) \mathbf{G}_m c_m(\boldsymbol{\theta}_m)}}{\sum_{m=1}^M \alpha_m \frac{c_m^T(\boldsymbol{\theta}_m) \mathbf{G}_m^T \mathbf{M}_m(\boldsymbol{\theta}_m, \mathbf{w}_{next}) \mathbf{G}_m c_m(\boldsymbol{\theta}_m)}{c_m^T(\boldsymbol{\theta}_m) \mathbf{G}_m c_m(\boldsymbol{\theta}_m)}} \right]^{-1}, \end{aligned}$$

where we have used the inequality

$$[1 + (1 - \gamma)(A - B)]^{-1} \geq \left[ \frac{A}{B} \right]^{-1}$$

for  $A \geq B \geq 0$ ,  $(1 - \gamma)B \leq 1$  and standard arguments in design theory.  $\square$

## References

- ATKINSON, A. C., DONEV, A. N. & TOBIAS, R. D. (2007) *Optimum Experimental Design, with SAS*, Oxford University Press, Oxford.
- ATKINSON, A. C. & FEDOROV, V. V. (1975) Optimal design: Experiments for discriminating between several models, *Biometrika* **62**, 289–303.
- BALTAGI, B. H. (2008) *Econometrics*, 4th edition, Springer, Berlin.
- BERGER, J. & PERICCHI, L. (1996) The intrinsic Bayes factor for model selection and prediction, *Journal of the American Statistical Association* **91**, 109–122.
- BIEDERMANN, S., DETTE, H. & PEPELYSHEV, A. (2006) Some robust designs for percentile estimation in binary response models, *Canadian Journal of Statistics* **34**, 603–622.
- BORNKAMP, B. (2006) Comparison of model-based and model-free approaches for the analysis of dose-response studies, Diploma Thesis, Fakultät Statistik, Technische Universität Dortmund, available under: <http://www.statistik.uni-dortmund.de/~bornkamp/diplom.pdf>.
- BORNKAMP, B., PINHEIRO, J. & BRETZ, F. (2010) *DoseFinding: Planning and Analyzing Dose Finding experiments*, R package version 0.1-1.
- BRETZ, F., HSU, J. C., PINHEIRO, J. C. & LIU, Y. (2008) Dose finding - a challenge in statistics, *Biometrical Journal* **50**, 480–504.
- BRETZ, F., PINHEIRO, J. & BRANSON, M. (2005) Combining multiple comparisons and modeling techniques in dose-response studies, *Biometrics* **61**(3), 738–748.
- CHALONER, K. & VERDINELLI, I. (1995) Bayesian experimental design: A review, *Statistical Science* **10**, 273–304.
- CHATFIELD, C. (1995) Model uncertainty, data mining and statistical inference, *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **158**, 419–466.

- COOK, R. D. & WONG, W. K. (1994) On the equivalence of constrained and compound optimal designs, *Journal of the American Statistical Association* **89**, 687–692.
- DETTE, H. (1996) Lower bounds for efficiencies with applications, in E. Brunner & M. Denker (eds.), *Research Developments in Probability and Statistics: Festschrift zum 65sten Geburtstag von M.L. Puri*, VSP Utrecht, pp. 111–124.
- DETTE, H., BRETZ, F., PEPELYSHEV, A. & PINHEIRO, J. C. (2008) Optimal designs for dose finding studies, *Journal of the American Statistical Association* **103**, 1225–1237.
- DETTE, H., KISS, C., BEVANDA, M. & BRETZ, F. (2010) Optimal designs for the emax, log-linear and exponential models, *Biometrika* , (in press).
- DETTE, H. & TITOFF, S. (2009) Optimal discrimination designs, *Annals of Statistics* **37**, 2056–2082.
- DRAGALIN, V., BORNKAMP, B., BRETZ, F., MILLER, F., PADMANABHAN, S. K., PATEL, N., PEREVOZSKAYA, I., PINHEIRO, J. & SMITH, J. R. (2010) A simulation study to compare new adaptive dose-ranging designs, *Statistics in Biopharmaceutical Research* , (in press).
- DRAGALIN, V., HSUAN, F. & PADMANABHAN, S. K. (2007) Adaptive designs for dose-finding studies based on sigmoid emax model, *Journal of Biopharmaceutical Statistics* **17**, 1051–1070.
- DRAPER, D. (1995) Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 45–97.
- FANG, K.-T. & WANG, Y. (1994) *Number theoretic methods in statistics*, Chapman and Hall, London.
- FEDOROV, V. & LEONOV, S. (2001) Optimal design of dose response experiments: A model-oriented approach, *Drug Information Journal* **35**, 1373–1383.
- GOLUB, G. & PEREYRA, V. (2003) Separable nonlinear least squares: the variable projection method and its applications, *Inverse Problems* **19**, R1–R26.

- HJORTH, J. S. U. (1994) *Computer intensive statistical methods: validation model selection and bootstrap*, Chapman and Hall, New York.
- KASS, R. & RAFTERY, A. (1995) Bayes factors, *Journal of the American Statistical Association* **90**, 773–795.
- KING, J. & WONG, W. K. (2004) Optimal designs for the power logistic model, *Journal of Statistical Computation and Simulation* **74**, 779–791.
- LÄUTER, E. (1974) Experimental design in a class of models, *Mathematische Operationsforschung und Statistik* **5**, 379–398.
- MILLER, F., GUILBAUD, O. & DETTE, H. (2007) Optimal designs for estimating the interesting part of a dose-effect curve, *Journal of Biopharmaceutical Statistics* **17**, 1097–1115.
- MÜLLER, P., BERRY, D. A., GRIEVE, A. P. & KRAMS, M. (2006) A Bayesian decision-theoretic dose-finding trial, *Decision Analysis* **3**, 197–207.
- O’HAGAN, A. (1995) Fractional Bayes factors for model comparison (with discussion), *Journal of the Royal Statistical Society B* **56**, 99–138.
- O’HAGAN, A. & FORSTER, J. (2004) *Kendall’s Advanced Theory of Statistics, Volume 2B: Bayesian Inference*, 2nd edition, Arnold, London.
- PINHEIRO, J. C., BORNKAMP, B. & BRETZ, F. (2006) Design and analysis of dose finding studies combining multiple comparisons and modeling procedures, *Journal of Biopharmaceutical Statistics* **16**, 639–656.
- PUKELSHEIM, F. (1993) *Optimal Design of Experiments*, Wiley, New York.
- RUBERG, S. J. (1995) Dose response studies. I. Some design considerations, *Journal of Biopharmaceutical Statistics* **5**(1), 1–14.
- SEBER, G. & WILD, C. (1989) *Nonlinear Regression*, Wiley, New York.
- VAN DER VAART, A. (1998) *Asymptotic Statistics*, Cambridge University Press.

- WATHEN, J. K. & THALL, P. F. (2008) Bayesian adaptive model selection for optimizing group sequential clinical trials, *Statistics in Medicine* **27**, 5586–5604.
- WU, C. F. J. (1988) Optimal design for percentile estimation of a quantal response curve, *in Optimal Design and Analysis of Experiments*, Elsevier, pp. 213–222.
- ZHOU, X., JOSEPH, L., WOLFSON, D. B. & BÉLISLE, P. (2003) A Bayesian A-optimal and model robust design criterion, *Biometrics* **59**, 1082–1088.
- ZHU, W. & WONG, W. K. (2000) Multiple-objective designs in a dose-response experiment, *Journal of Biopharmaceutical Statistics* **10**, 1–14.
- ZHU, W. & WONG, W. K. (2001) Bayesian optimal designs for estimating a set of symmetric quantiles, *Statistics in Medicine* **20**, 123–137.